

Model-Based Data Interpretation and Diagnosis Robustness

R. Pasquier

*Applied Computing and Mechanics Laboratory (IMAC),
School of Architecture, Civil and Environmental Engineering (ENAC),
École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

J.-A. Goulet

*Department of Civil and Environmental Engineering
University of California, Berkeley, USA*

I.F.C. Smith

*Applied Computing and Mechanics Laboratory (IMAC),
School of Architecture, Civil and Environmental Engineering (ENAC),
École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

ABSTRACT: Model-based data-interpretation techniques are widely used to identify the behavior of structures. These techniques exploit information provided by in-situ measurements to make diagnosis related to structural performance. In the context of model-based infrastructure diagnoses, measuring and modeling uncertainties are generally estimated using engineering heuristics. The accuracy of diagnosis is related to the accuracy of uncertainty estimation. For civil infrastructure diagnoses, the estimation of modeling uncertainties is non-trivial, because special attention is required to avoid diagnosis errors. For data interpretation methodologies that generate multiple candidate models, a diagnosis error occurs when *incorrect models are accepted, while the correct model is rejected*. The probability of diagnosis error is sensitive to two factors: (1) misevaluation of uncertainties and (2) the number of measurements used for data interpretation. In this context, the robustness is defined as the ability of providing the right diagnosis in presence of misevaluation of uncertainties. This paper presents a preliminary study that quantifies the sensitivity of diagnosis to errors with respect to misevaluation of uncertainties and the number of measurements used. The study of a beam example shows that when the mean of uncertainty is misestimated, the probability of diagnosis error increases with the number of measurements. Inversely, when the uncertainty standard deviation is underestimated, the probability of diagnosis error decreases with the number of measurements. For the case where both uncertainty mean and standard deviation are misevaluated, it is possible to find a minimum number of measurements that assures the diagnosis robustness.

1 INTRODUCTION

Model-based data-interpretation techniques are widely used to identify the behavior of structures. These techniques exploit information provided by in-situ measurements to make diagnosis related to structural performance. Measured data are used to infer unknown parameter values of behavior models. Since data interpretation of civil structures is abductive, multiple models explain the true structural behavior. Except in unusual circumstances, the presence of uncertainties increase the number of these models.

Many diagnosis methodologies that are able to handle multiple models are available. One of the most widely used is Bayesian model updating (Beck and

Katafygiotis 1998, Kennedy and O'Hagan 2001, Yuen 2010). This approach uses prior knowledge of the model parameters and a likelihood function to infer parameter values based on observations. Beven and Binley (1992) proposed the Generalized Likelihood Uncertainty Estimation (GLUE) to identify the properties of system when little information is available to define uncertainties. This approach was developed in the context of environmental applications where uncertainties are high, and where it is difficult to assign probabilistic models to these uncertainties. Goulet et al. (2012) proposed a probabilistic model-falsification methodology called error-domain model falsification (EDMF). This approach is similar to the GLUE approach except that it compares the difference between predicted and

measured values with threshold values that are defined probabilistically.

In the context of model-based infrastructure diagnoses, measuring and modeling uncertainties are generally estimated using engineering heuristics. For Bayesian model updating, a common assumption is to define these uncertainties as zero-mean Gaussian white noise (Katafygiotis et al. 1998, Beck and Yuen 2004, Goller and Schuëller 2011). When the error structure is known, Bayesian approaches are able to handle uncertainties as unknown parameters during data interpretation. Recent studies considered more complex probabilistic models to describe modeling errors (Papadimitriou and Lombaert 2012, Simoen et al. 2013). Applications of the EDMF methodology included both measuring and modeling uncertainties (Goulet et al. 2012, Goulet and Smith 2012, Goulet et al. 2013). These uncertainties were either estimated based on statistical procedures or they were estimated using engineering heuristics. For every approach, the accuracy of results is related to the accuracy of uncertainty estimation. Moreover, the more measurements are used for data interpretation, the larger is the chance of committing a misevaluation of one of these uncertainties. For civil infrastructure diagnoses, the estimation of modeling uncertainties is non-trivial, and therefore special attention is needed to either avoid diagnosis errors or devise interpretation methodologies that are robust in their presence.

In the field of statistics, hypothesis testing is used for testing the trueness of a statement. The null hypothesis is defined as the true statement and, it is either accepted or rejected based on an approximating distribution of the parameter subjected to hypothesis. Here, because the parameter estimation is based on random samples, diagnosis errors are inevitable (Miller and Freund 1977). A false rejection of the null hypothesis is called type I error and a false acceptance of an alternative hypothesis is defined as the type II error. In the context of civil engineering and solid mechanics, hypothesis testing was used to test the validity of models including the consideration of the probability of type I and type II errors (Rebba and Mahadevan 2006, McFarland and Mahadevan 2008, Jiang and Mahadevan 2008).

By extension, for data interpretation methodologies handling multiple models, type I error is the rejection of the correct model and type II error is the acceptance of one or several incorrect models in addition to the correct model. There is a diagnosis error when *incorrect models are accepted, while the correct model is rejected*. Thus, diagnosis error is defined as the simultaneous occurrence of type I and type II errors. The probability of diagnosis error is sensitive to two factors: (1) misevaluation of uncertainties and (2) the number of measurements used for data interpretation. In this context, the robustness is defined as the ability of providing the right diagnosis in presence of misevaluation of uncertainties. Moreover, the diagnosis

robustness is inversely related to diagnosis sensitivity with respect to misevaluation of uncertainties. The study of diagnosis sensitivity with respect to misevaluation of uncertainties may reveal the minimum number of measurements for which diagnosis robustness is guaranteed. Ben-Haim and Hemez (2012) studied the robustness of models based on the comparison of model predictions and test data. They demonstrated analytically that improving the model robustness to incomplete understanding of the physical process (i.e. modeling uncertainties) decreases the fidelity (i.e. low discrepancy) of predictions to test data. Indeed, the fidelity of predictions to test data is closely related to the false-rejection and false-acceptance rates during data interpretation. By extension, the probability of type I and type II errors is related to robustness. The effect of misevaluation of uncertainties and additional measurements on the robustness was, however, not studied.

This paper presents a preliminary study that quantifies the sensitivity of diagnosis to errors with respect to misevaluation of uncertainties and the number of measurements used. This study is made in the context of the error-domain model falsification methodology. A summary of this methodology is presented in Section 2. Section 3 describes the methodology proposed in this paper to evaluate the probability of committing diagnosis errors with respect to the number of measurements used, and to the nature of uncertainty misevaluation. Section 4 presents the results of the preliminary study including three scenarios examined for a cantilever-beam example.

2 MODEL-BASED DATA INTERPRETATION

In this paper, the methodology used for data-interpretation is error-domain model falsification. In order to predict the behavior of a structure studied, a model class $g(\theta)$ is created. This model class takes as input n_p physical parameters $\theta = [\theta_1, \theta_2, \dots, \theta_{n_p}]^T$ describing the unknown characteristics of the structure. The domain of the physical parameters is explored by generating an initial set of *model instances* corresponding to discrete sets of parameter values evaluated using a model class. Error-domain model falsification is based on the comparison of predicted and measured values such as

$$g_i(\theta) - y_i = \epsilon_{model,i} - \epsilon_{measure,i} \quad (1)$$

where y_i is the measured value for the comparison point i , and where $\epsilon_{model,i}$ and $\epsilon_{measure,i}$ are the model and measurement errors for the comparison point i . Both the model error $\epsilon_{model,i}$ and measurement error $\epsilon_{measure,i}$, are represented by random variables $U_{model,i}$ and $U_{measure,i}$. Model uncertainties are, for example, due to the geometric variability of the structure, the variability of material properties, the mesh-refinement

and, uncertainties associated with model simplifications. Model and measurement uncertainties are combined such as

$$U_{c,i} = U_{model,i} - U_{measure,i}$$

where $U_{c,i}$ represents the possible outcomes of the differences between predicted and measured values. A model instance is falsified if, for any of the n_m comparison points, the difference between predicted and measured values is outside the interval defined by threshold bounds (see Equation 2). A model instance is a candidate model only if

$$\forall i \in \{1, \dots, n_m\} : T_{low,i} \leq g_i(\theta) - y_i \leq T_{high,i} \quad (2)$$

i.e. the difference between predictions and measurements lies inside the intervals defined by threshold values for all i (measurement locations).

Threshold values are determined for every quantity i based on the combined uncertainty probability density function $f_{U_c}(\epsilon_c)$. Threshold values $T_{low,i}$ and $T_{high,i}$ define the shortest intervals satisfying

$$\phi^{1/n_m} = \int_{T_{low,i}}^{T_{high,i}} f_{U_{c,i}}(\epsilon_{c,i}) d\epsilon_{c,i} \quad \forall i \in \{1, \dots, n_m\} \quad (3)$$

where $\phi \in]0, 1]$ is a target probability defined by users. The value of ϕ is the target probability that correct models remain in the candidate model set after falsification. This choice for threshold values uses the Šidák correction to include the effects of the comparison of multiple predictions and measurements (Šidák 1967). It has the advantage of providing conservative threshold values, regardless of the dependencies between uncertainties (JCGM 2011).

All models that are not falsified are kept in what is called the *candidate model set*. This set contains model instances that are compatible with measurements while given model and measurement uncertainties.

3 DIAGNOSIS-ERROR SENSITIVITY

The methodology proposed in this paper quantifies the sensitivity of diagnosis to misevaluating uncertainties and to the number of measurements used for data interpretation. It uses the error-domain model falsification methodology presented in Section 2. The term model and model instance are used as synonyms for instantiation of parameters within a model class. When uncertainties are misestimated, it is possible to accept incorrect models while falsely rejecting correct models. Minimizing the risk of such scenarios requires knowledge of two factors: the effect of misevaluation of uncertainties and the effect of the number of measurements. These factors produce two competing effects;

(1) when modeling uncertainties are misestimated for several predictions, increasing the number of measurements, increases the difficulty of not falsifying correct models. Thus, the probability of having false-rejection errors $P(\alpha)$ (i.e. type I error) increases with the number of measurements used. (2) The larger the number of measurements, the lower the probability of accepting incorrect models $P(\beta)$ (i.e. type II error). The simultaneous occurrence of the false-acceptance and the false-rejection event represents the probability of committing a diagnosis error $P(\alpha \cap \beta)$. Note that the null hypothesis is the occurrence of a model being accepted as a candidate model.

Diagnosis robustness is evaluated by studying the sensitivity of $P(\alpha \cap \beta)$ to misevaluation of uncertainties and to the number of measurements. This study uses simulated measurements to emulate the process of model falsification. Simulated measurements \hat{Y}_i are treated as random variables obtained from

$$\hat{Y}_i = g_i(\theta^*) - \tilde{U}_i \cdot g_i(\bar{\theta}), \forall i \in \{1, \dots, n_m\} \quad (4)$$

where θ^* is a ad-hoc defined true parameter set, and U_i is a random error for each of the n_m measurement locations. The error \tilde{U}_i is multiplied by the prediction values of the model class evaluated with the mean value of the parameter set $\bar{\theta}$. This operation aims to separate errors that depend on the model from errors that are due to the variability of the parameter values θ , used for the generation of simulated measurements. The modeling error \tilde{U}_i presented in Eq.(4) is estimated by a Gaussian uncertainty distribution ($\tilde{U}_i \sim \mathcal{N}(\mu_0 + \Delta\mu, (\sigma_0 + \Delta\sigma)^2)$). In order to represent the situation where uncertainties are misevaluated, a bias is added to the mean ($\Delta\mu$) and the standard deviation ($\Delta\sigma$) during the generation of simulated measurements. Mean and standard deviation values used in the data-interpretation approach (i.e. to define the threshold bounds) are μ_0 and σ_0 .

The relative frequency of simultaneous occurrence of type I ($P(\alpha)$) and type II ($P(\beta)$) errors is used to quantify the sensitivity to diagnosis error $P(\alpha \cap \beta)$. In order to quantify $P(\alpha \cap \beta)$, N_{SM} sets of simulated measurements are generated. At each step, model instances that are not compatible with simulated measurements are falsified. The *false-rejection* probability is

$$P(\alpha) = N_\alpha / N_{SM}$$

This is the ratio between the number of times that the correct model is wrongly falsified N_α , and the number of sets of simulated measurements N_{SM} . The probability of *incorrect-model acceptance* is

$$P(\beta) = N_\beta / N_{SM}$$

$P(\beta)$ is computed by dividing the number of times that incorrect models are accepted N_β , by N_{SM} . The probability of type I and type II errors is obtained

from the relative number of simultaneous occurrence of false rejection with false acceptance $N_{\alpha\cap\beta}$ such that

$$P(\alpha \cap \beta) = N_{\alpha\cap\beta}/N_{SM}$$

where $P(\alpha \cap \beta)$ is the *probability of diagnosis error*. An example of diagnosis error is illustrated in Figure 1. This figure presents the measured value and the predicted values of every model instance. Candidate models are found inside the threshold bounds, whilst the correct model is falsified, leading to a diagnosis error.

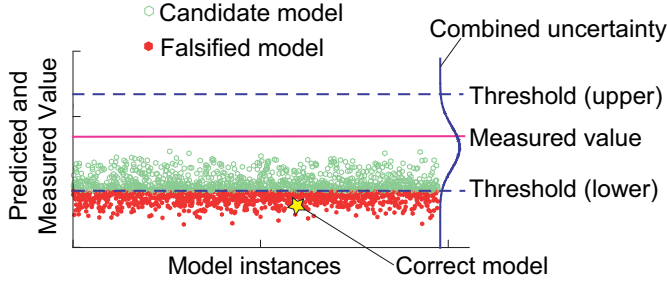


Figure 1: Diagnosis-error example in the error domain. The figure shows a false acceptance of candidate models along with false rejection of the correct model.

4 ILLUSTRATIVE EXAMPLE

The diagnosis robustness is studied for a cantilever-beam example depicted in Figure 2. The beam is $l = 3000$ mm long and the inertia of its square cross-section is $I = 6.75 \times 10^8$ mm⁴, for $h = 300$ mm.

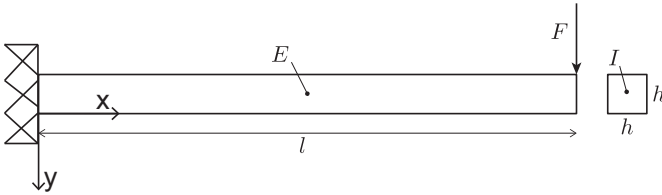


Figure 2: Cantilever-beam example.

The system has two unknown parameters to be identified: the Young's modulus E and the force F applied at the right end of the beam ($\theta = [E, F]^T$). The model $g(\theta)$ used to predict the vertical displacements is

$$g(x, \theta) = \frac{Fx^2(3l - x)}{6EI} \quad (5)$$

An initial population of models containing 10 000 instances is generated. Model instances are evaluated at the intersection of a grid where the Young's modulus is discretized in intervals of 0.8 GPa over the range $[20 - 100]$ GPa and the applied force is discretized in intervals of 0.09 kN over the range $[1 - 10]$ kN. In this example, the target probability ϕ is set to 0.95. The

diagnosis robustness is evaluated for several values of number of measurements n_m . Measurement locations are uniformly distributed between the coordinate $x = x_{\text{start}} = 500$ mm and $x = l$. The coordinate x_i of each sensor is defined by

$$x_i = x_{\text{start}} + i \cdot \frac{l - x_{\text{start}}}{n_m}, \quad \forall i \in \{1, \dots, n_m\} \quad (6)$$

The model predictions $g(\theta)$ are affected by errors that are assumed to be Gaussian with zero mean and a standard deviation corresponding to 5% of the predicted value (i.e. $U_{c,i} \sim \mathcal{N}(0, (0.05 \cdot g_i(\bar{\theta}))^2)$).

The probabilities $P(\alpha)$, $P(\beta)$ and $P(\alpha \cap \beta)$ are studied for several misvaluations of mean and standard deviation values. In a first scenario, $\Delta\sigma = 0$ and a bias $\Delta\mu \geq 0$ is added to the mean value of the modeling uncertainty to simulate the effect of uncertainty misvaluation. In a second scenario, a bias $\Delta\sigma \geq 0$ is added to the standard deviation of the modeling uncertainty and $\Delta\mu = 0$. In a third scenario, $\Delta\mu$ and $\Delta\sigma$ are generated stochastically. $\Delta\mu$ follows a Gaussian distribution and $\Delta\sigma$ a Gamma distribution. The statistical parameters of these distribution are selected in relation with the results found for the first and second scenarios. Scenarios I to III are summarized in Table 1.

Table 1: Summary of scenarios.

Scenario	Analysis type	$\Delta\mu$	$\Delta\sigma$
I	deterministic	≥ 0	$= 0$
II	deterministic	$= 0$	≥ 0
III	stochastic	$\sim \mathcal{N}(\cdot)$	$\sim \Gamma(\cdot)$

4.1 Scenario I

In the first scenario, the bias on the uncertainty mean value is equal for each measurement location. Values $\Delta\mu = \{0, 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.25, 0.5, 5\}$ are used to evaluate the diagnosis sensitivity to type I and type II errors. Figure 3 shows a schematic representation of a set of two simulated measurements for this scenario. In this figure, both model predictions $g_i(\theta^*)$, with $i = \{1, 2\}$, are represented on the bottom face of the deformed cantilever beam. For each measurement location i , an error $\tilde{U}_i \cdot g_i(\bar{\theta})$, which is randomly selected from the biased uncertainty distribution, is subtracted from the model prediction to obtain the simulated measurement \hat{Y}_i . The 0-value of the uncertainty distribution is coincident with the bottom face of the beam and the mean value of the uncertainty is biased to $\Delta\mu \cdot g_i(\bar{\theta})$.

Figure 4a presents the probability of type I error ($P(\alpha)$), as well as the probability of type II error ($P(\beta)$), with respect to the number of measurements used for data interpretation and the bias value for $\Delta\mu$. Figure 4b illustrates the simultaneous occurrence of

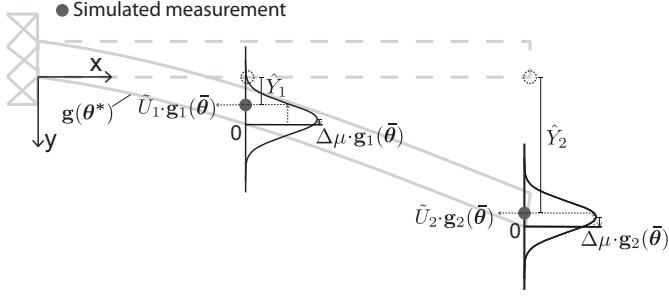


Figure 3: Simulated measurements emulating a misevaluation of the modeling uncertainty mean.

both errors ($P(\alpha \cap \beta)$) with respect to the number of measurements and the bias value for $\Delta\mu$. These results are compared with the case without uncertainty misevaluation (i.e. $\Delta\mu = 0$ and $\Delta\sigma = 0$) where the probability of simultaneous occurrence of both errors is at most $1 - \phi = 0.05$ regardless of the number of measurements.

In Figure 4a, the probability of type I error increases up to 1 when the number of measurements increases. The higher the bias, the lower the number of measurements needed to reach $P(\alpha) = 1$. For the probability of type II error, the higher the bias, the lower $P(\beta)$, the probability remaining close to 1 for any value of number of measurements with $\Delta\mu < 0.25$. If the bias is large (e.g. $\Delta\mu = 5$), all model instances are always falsified and $P(\beta) = 0$.

In Figure 4b, $P(\alpha \cap \beta) \approx P(\alpha)$ for $\Delta\mu < 0.25$. When $\Delta\mu \geq 0.25$, $P(\alpha \cap \beta) \approx P(\beta)$ is nearly constant regardless of the number of measurements. For these values, the bias is so high that $P(\alpha) = 1$ with any value of number of measurements. Thus, $P(\alpha \cap \beta)$ is governed by the number of false acceptances. With $\Delta\mu = 5$, the entire population of models is falsified for any value of number of measurements.

$P(\beta)$ is less sensitive to the changes in the value of number of measurements than $P(\alpha)$, because for any $\Delta\mu$, the deformed shape of model instances are analogous to the deformed shape of the simulated measurements. Figure 5 presents a comparison of ten sets of five simulated measurements with the predictions of the least and most rigid models. According to Eq.(4), increasing the uncertainty mean value of the uncertainty distribution, shifts the mean of the simulated measurements. Because the random sample of \tilde{U}_i is subtracted from the model prediction $g_i(\theta^*)$ for each measurement location, the simulated measurements are shifted all in the same direction compared with the model prediction. The geometry drawn by the set of simulated measurements is analogous to the model deformed shape, returning a frequent acceptance of incorrect models. Thus, even by adding new measurements, the probability of false acceptance remains about constant.

As seen in this scenario, the diagnosis is sensitive to misevaluation of the uncertainty mean value. Diagnosis sensitivity to the number of measurements

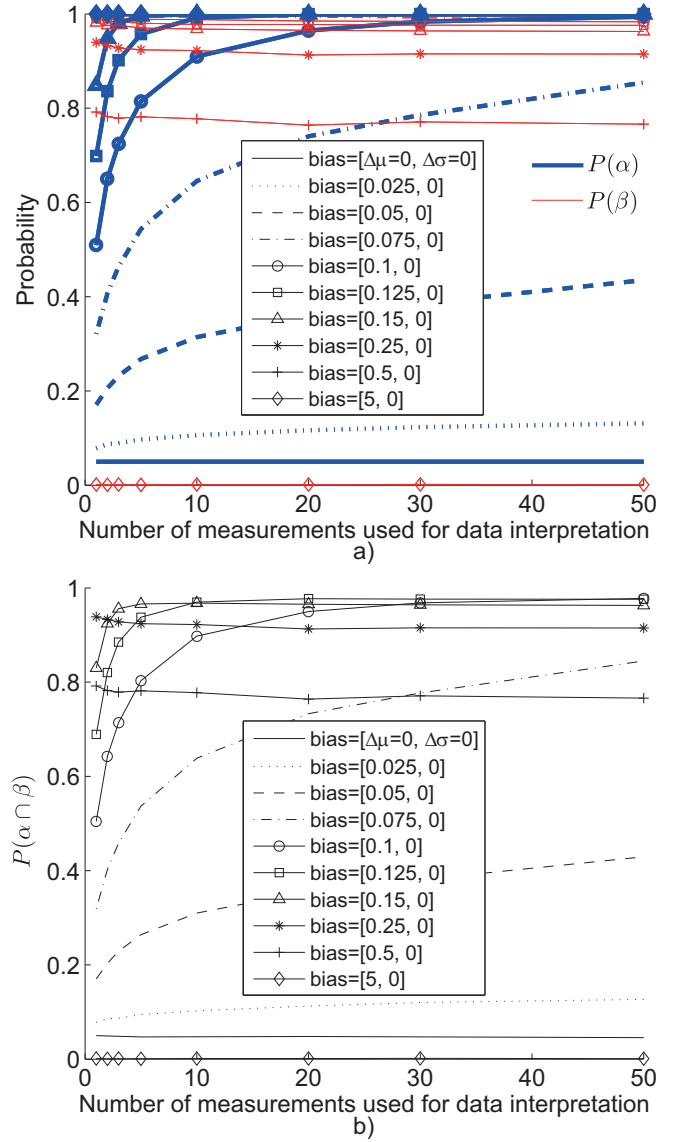


Figure 4: Probability of committing a diagnosis error with respect to the number of measurements used for data interpretation for different bias values $\Delta\mu$. a) Type I error $P(\alpha)$ and type II error $P(\beta)$. b) Combination of both type I and type II errors $P(\alpha \cap \beta)$.

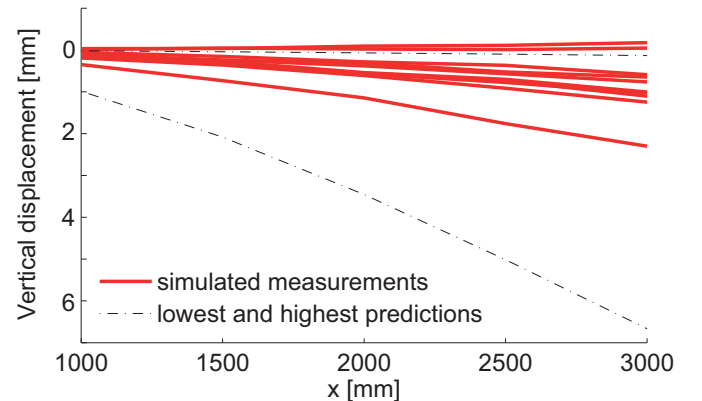


Figure 5: Examples of simulated measurements in comparison with the lowest and highest model predictions in scenario I.

is high for bias values $\Delta\mu < 0.25$. The diagnosis is not sensitive to the number of measurements for bias values $\Delta\mu \geq 0.25$. Thus, increasing the number of measurements does not improve diagnosis robustness for $\Delta\mu \geq 0.25$ and even deteriorates the robustness for $\Delta\mu < 0.25$.

4.2 Scenario II

In the second scenario, the bias on the standard deviation value is also equal for each measurement location and $\Delta\sigma = \{0, 0.005, 0.025, 0.05, 0.1, 0.25, 0.5\}$. Figure 6 represents schematically the generation of a set of two simulated measurements for this scenario. In this figure, both model predictions $g_i(\theta^*)$, with $i = \{1, 2\}$, are represented on the bottom face of the deformed cantilever beam. For each measurement location i , an error $\tilde{U}_i \cdot g_i(\bar{\theta})$, which is randomly selected from the biased uncertainty distribution, is subtracted from the model prediction to obtain the simulated measurement \hat{Y}_i . The 0-value of the uncertainty distribution is coincident with the bottom face of the beam and the standard deviation value of the uncertainty is biased to $\Delta\sigma \cdot g_i(\bar{\theta})$.

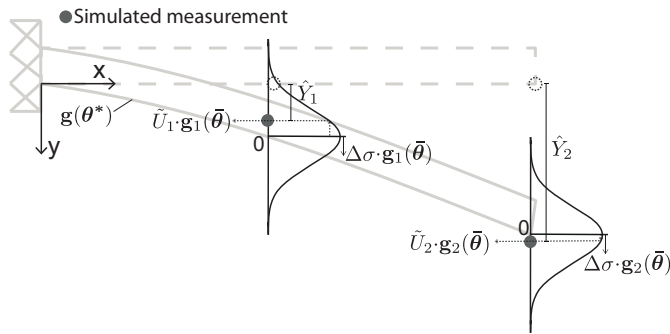


Figure 6: Simulated measurements emulating a misvaluation of the modeling uncertainty standard deviation.

As in the first scenario, Figure 7a presents the probability of type I error ($P(\alpha)$), as well as the probability of type II error ($P(\beta)$), with respect to the number of measurements and the bias value for $\Delta\sigma$. Figure 7b illustrates the simultaneous occurrence of both errors ($P(\alpha \cap \beta)$) with respect to the number of measurements and the bias value for $\Delta\sigma$.

Figure 7a shows that the probability $P(\alpha)$ increases up to 1 with the number of measurements. The higher $\Delta\sigma$, the lower the number of measurements needed to reach $P(\alpha) = 1$. Contrarily to the scenario I, $P(\beta)$ decreases to 0 with the number of measurements. Also, the higher $\Delta\sigma$, the lower the number of measurements needed to reach $P(\beta) = 0$.

Figure 7b shows the simultaneous occurrence of type I and type II errors, $P(\alpha \cap \beta)$. For $\Delta\sigma < 0.1$, the probability of diagnosis error increases until reaching a maximum and then decreases slightly when new measurements are added. For $\Delta\sigma \geq 0.1$, $P(\alpha \cap \beta)$ decreases with the addition of measurements. The curves

decrease steeply to $P(\alpha \cap \beta) = 0$, indicating that 100% of the models are falsified. The higher the bias, the steeper is the reduction in $P(\alpha \cap \beta)$.

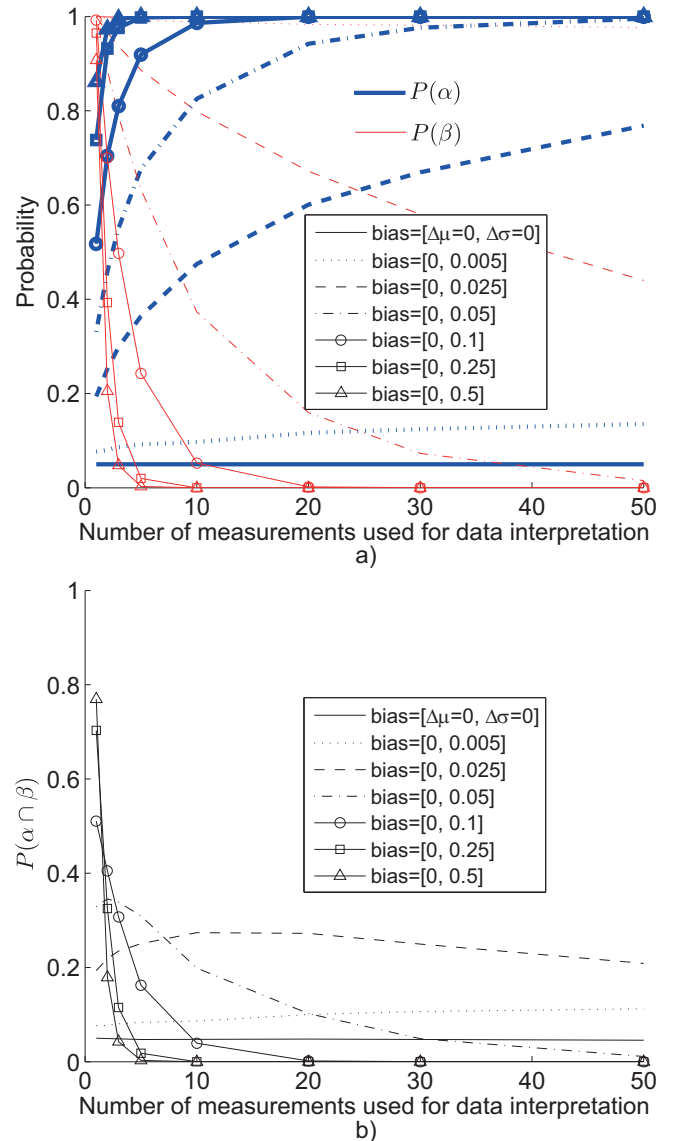


Figure 7: Probability of committing a diagnosis error with respect to the number of measurements used for data interpretation for different bias values $\Delta\sigma$. a) Type I error $P(\alpha)$ and type II error $P(\beta)$. b) Combination of both type I and type II errors $P(\alpha \cap \beta)$.

Figure 8 compares ten sets of five simulated measurements with the predictions of the softest and stiffest models. According to Eq.(4), increasing the standard deviation value of the uncertainty distribution raises the dispersion of the simulated measurements. Thus, because of the random selection of the sample \tilde{U}_i at each measurement location, each simulated measurement is deviated on both sides of the model-prediction value $g_i(\theta^*)$. Therefore, the geometry drawn by simulated measurements is not compatible with model predictions. This limits the occurrence of a false acceptance. By adding new measurements, the probability that the geometry becomes not compatible with model predictions increases and $P(\beta)$ decreases.

In this scenario, the diagnosis sensitivity to mis-

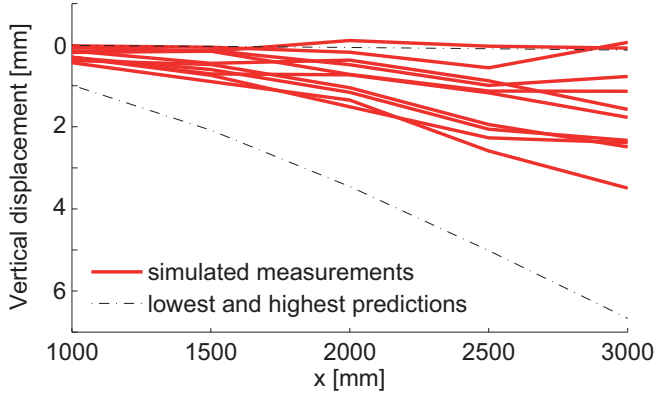


Figure 8: Examples of simulated measurements in comparison with the lowest and highest model predictions in scenario II.

evaluation of the standard deviation value increases when the bias value increases for a single measurement. However, this sensitivity decreases greatly by adding new measurements for $\Delta\sigma \geq 0.1$. Thus, adding new measurements leads to an improvement of diagnosis robustness. Moreover, when $P(\alpha \cap \beta) = 0$, the diagnosis robustness is assured.

4.3 Discussion of scenarios I and II

In the first two scenarios, the diagnosis robustness is studied with biases selected using engineering heuristics. The results obtained with this simplified representation provides a good understanding of the mechanisms leading to diagnosis error in case of misestimation of uncertainty mean and standard deviation during data interpretation. For equivalent bias values, the results show that the diagnosis sensitivity to uncertainty misvaluation is higher when the mean value is misestimated than when the standard deviation is underestimated. Moreover, in the second scenario, for a defined value of number of measurements, diagnosis robustness is assured when the entire population of models is falsified ($P(\alpha \cap \beta) = 0$). Contrarily to the first scenario, adding new measurements has a negative effect on diagnosis robustness.

A limitation of these two scenarios is that they do not consider different bias values for each comparison point (measurement location). This aspect is studied in the third scenario.

4.4 Scenario III

When interpreting real data at several comparison points, the chance of misvaluing uncertainties is not perfectly correlated between the comparison points as illustrated in scenarios I and II. While the chance of misvaluing uncertainties is not independent either, the reality is situated between these two extremes. In addition, the amplitude of the bias value is not constant for

all comparison points. Thus, scenario III reflects better the reality than the previous scenarios. Biases for $\Delta\mu$ are selected using a Gaussian distribution $\mathcal{N}(0, 6.25)$ and biases for $\Delta\sigma$ are obtained from a Gamma distribution of $\Delta\sigma^2$. Both distributions are presented in Figure 9. The parameters of these distributions are chosen such that $\Delta\mu$ has 5% probability to be higher than 5 and $\Delta\sigma$ has 5% probability to be higher than 0.25. As seen in scenarios I and II, for higher bias values than these ones, the probability of diagnosis error reaches $P(\alpha \cap \beta) = 0$ when the number of measurement is large. For each simulated measurement, values for $\Delta\mu$ and $\Delta\sigma$ are randomly generated and added to the uncertainty mean and standard deviation values as described in Eq.(4). Therefore, bias values are different for each comparison point.

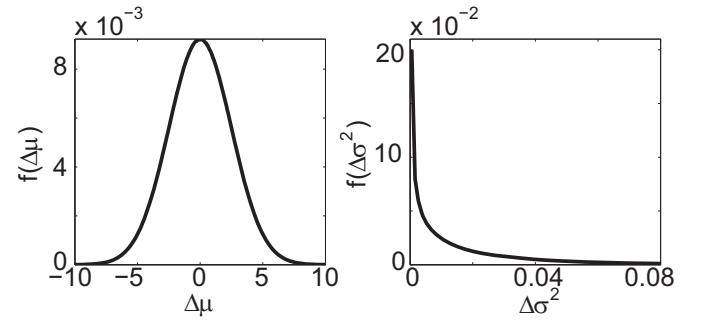


Figure 9: Gaussian distribution of $\Delta\mu$ and Γ distribution of $\Delta\sigma^2$.

Figure 10 presents the probability of diagnosis error with respect to the number of measurements for the third scenario. Using a single measurement leads to a probability of 59% of committing a diagnosis error, while using two measurements lead to 4% probability of diagnosis error. Using more than three measurements lead to the falsification of the entire initial model set and $P(\alpha \cap \beta) = 0$. For these values of number of measurements, misvaluation of uncertainties does not affect the probability of diagnosis error.

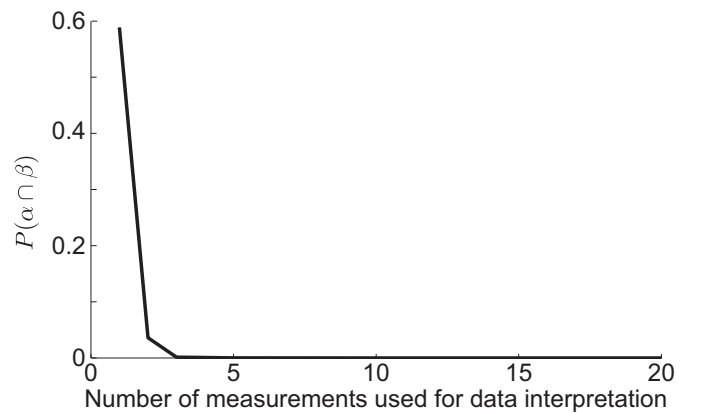


Figure 10: Probability of committing a diagnosis error using stochastic biases.

When the uncertainty misvaluation is different for each comparison point, the diagnosis is highly sensitive to the number of measurements. This is caused

by the same reasons mentioned in scenario II and illustrated by Figure 8. Here, the increasing dispersion of the simulated measurements is due to the variance of the uncertainty as well as the random positive and negative sign that takes $\Delta\mu$.

These results indicate that for this scenario, at least three measurements should be used to guarantee the diagnosis robustness. In case of uncertainty misevaluation, the data interpretation would at worst return no diagnosis (i.e. falsification of the entire population of models). This indicates that either the model class that is selected for data interpretation is incorrect or the modeling error is misevaluated. In this simple example, only the second proposition is true, while in real applications both are possible. In such situations, initial assumptions need to be revised in order to achieve a correct diagnosis.

5 CONCLUSION

This study found that the diagnosis errors are sensitive to the number of measurements used. The preliminary study of a beam example shows that:

- The probability of diagnosis errors increases with the number of measurements, when the uncertainty mean value is misestimated.
- The probability of diagnosis errors decreases with the number of measurements, when the uncertainty standard deviation is underestimated.
- For the case where both uncertainty mean and standard deviation are misevaluated, it is possible to find a minimum number of measurements that guarantees the diagnosis robustness (i.e. misevaluating uncertainties does not affect the probability of having diagnosis errors).

The results of the study of diagnosis sensitivity provide guidance for designing measurement systems that are robust with respect to diagnosis error. More effort is needed to adapt this methodology to full-scale civil infrastructure, where modeling errors may be systematic and correlated spatially as well as among quantities.

REFERENCES

- Beck, J. & L. Katafygiotis (1998). Updating models and their uncertainties. i: Bayesian statistical framework. *Journal of Engineering Mechanics* 124(4), 455–461.
- Beck, J. & K. Yuen (2004). Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics* 130(2), 192–203.
- Ben-Haim, Y. & F. Hemez (2012). Robustness, fidelity and prediction-looseness of models. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 468(2137), 227–244.
- Beven, K. & A. Binley (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* 6(3), 279–298.
- Goller, B. & G. Schuëller (2011). Investigation of model uncertainties in bayesian structural model updating. *Journal of sound and vibration* 330(25), 6122–6136.
- Goulet, J., C. Michel, & I. Smith (In press, 2012). Hybrid probabilities and error-domain structural identification using ambient vibration monitoring. *Mechanical Systems and Signal Processing*.
- Goulet, J. & I. Smith (2012). Predicting the usefulness of monitoring for identifying the behavior of structures. *Journal of Structural Engineering*.
- Goulet, J.-A., S. Coutu, & I. F. Smith (In press, 2013). Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Advanced Engineering Informatics*.
- JCGM (2011). *Evaluation of measurement data – Supplement 2 to the “Guide to the expression of uncertainty in measurement” – Extension to any number of output quantities*, Volume JCGM 102:2011. JCGM Working Group of the Expression of Uncertainty in Measurement.
- Jiang, X. & S. Mahadevan (2008). Bayesian validation assessment of multivariate computational models. *Journal of Applied Statistics* 35(1), 49–65.
- Katafygiotis, L., C. Papadimitriou, & H. Lam (1998). A probabilistic approach to structural model updating. *Soil Dynamics and Earthquake Engineering* 17(7), 495–507.
- Kennedy, M. C. & A. O’Hagan (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(3), 425–464.
- McFarland, J. & S. Mahadevan (2008). Multivariate significance testing and model calibration under uncertainty. *Computer Methods in Applied Mechanics and Engineering* 197(29-32), 2467–2479.
- Miller, I. & J. E. Freund (1977). *Probability and Statistics for Engineers* (2nd ed.). Prentice-Hall Inc.
- Papadimitriou, C. & G. Lombaert (2012). The effect of prediction error correlation on optimal sensor placement in structural dynamics. *Mechanical Systems and Signal Processing* 28, 105–127.
- Rebba, R. & S. Mahadevan (2006). Validation of models with multivariate output. *Reliability Engineering & System Safety* 91(8), 861–871.
- Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 62(318), 626–633.
- Simoen, E., C. Papadimitriou, G. De Roeck, & G. Lombaert (2013). Influence of the prediction error correlation model on bayesian fe model updating results. In *Life-Cycle and Sustainability of Civil Infrastructure Systems*, pp. 192–199.
- Yuen, K. (2010). *Bayesian methods for structural dynamics and civil engineering*. Wiley.
- R. Pasquier, J.-A. Goulet and I. F.C. Smith. Model-Based Data Interpretation and Diagnosis Robustness. Safety, Reliability, Risk and Life-Cycle Performance of Structures & Infrastructures. Proceedings of 11th International Conference on Structural Safety & Reliability (ICOSSAR), New York, USA, 2013, 2497-2504.